7-2017

Multi-Choice Questions and Their Problems When Used for Assessment of Aircraft Engineers Education

Ian R. McAndrew
Embry-Riddle Aeronautical University, mcand4f1@erau.edu

Ken L. Witcher
Embry-Riddle Aeronautical University, witchea8@erau.edu

Elena Navarro
Embry-Riddle Aeronautical University, navarrj1@erau.edu

Peter Foreman
Resource Group, Kemble, UK

Multi-Choice Questions and their Problems when used for assessment of Aircraft Engineers Education

Ian R. McAndrew PhD FRAeS[1]

Ken L. Witcher PhD FRAeS[2]

Elena Navarro MSCS[3]

Peter Foreman[4]

[1,2&3]Embry Riddle Aeronautical University – Worldwide – Daytona Beach

[4]Resource Group, Kemble, UK

Abstract

Licensed aircraft engineers under the European Aviation Safety Agency, EASA, undertake academic training to complement their practical and type specific studies. These exams are mainly Multi-Choice Questions, MCQ, and four 20-minute essays. The MCQ exams are as few as 16 questions to a maximum of 140 questions. A score of 75% is needed to pass each exam, and each question has three possible answers. This authors of this paper reviews the theory and design of the MCQ and asks if the assumptions are valid and that it achieves the academic level assumed for engineers who will be maintaining some of the most complex system in the world, and the safety of passengers. It will argue that there are failings and how this can be address, in particular, that repeated tests should have a higher pass level.

Keywords: Multi-Choice Questions, Assessment and Training.

Introduction

EASA took over the responsibilities of the Joint Aviation Authority, JAA, in the early part of the century, and while it has many responsibilities from Airworthiness of foreign carriers entering Europe to advising the European Union on legislation [1]. The principle focus in this paper is the certifying of licensed engineers to maintain aircraft [2]. There are other classification; however, this paper addresses those mainly concerned with commercial aircraft with a Mean-Take-Off-Weight, MTOW, greater than 5700 kg. Three classifications exist:

> Category A, where maintainers are allowed to undertake basic tasks unsupervised, other tasks if supervised but not unsupervised on safety critical aspects and are limited in what they are allowed to certify, and privileges are restricted to work that the maintainer has personally performed in a Part-145 organisation.
>
> Category B is divided in to Mechanical (B1) and Avionics (B2) and these are authorized to undertake complex tasks unsupervised and certify the work completed. They are the group that are able to undertake all work between their skill set.
>
> Category C is the highest level and they are authorized to sign off an aircraft after base maintenance has been completed and state it is safe to return to airworthy status.

All of these levels can only perform their functions if working for an approved part 145 organisation under part 66. In the United States, the equivalent is an A & P engineer (Airframe and Power plant). They can certify their own work under part 65 independently. Regardless that they are different in some ways the training of EASA or A&P is a serious task as lives are at risk if unsuitable or unqualified personnel are allowed to maintain aircraft [3].

There are set Exams for a variety of courses and Table 1 below summarizes the exam requirement for the three classic license levels and the number of questions in each exam. Time allocated for each question is 1 minute 15 seconds, for all questions, in all subjects and levels.

Module Title	**Module No.**	**A**	**B1**	**B2**	**Essay**
Mathematics	1	16	30	30	
Physics	2	30	52	52	
Electrical Fundamentals	3	20	52	52	
Electronic Fundamental	4	N/A	20	40	
Digital Techniques	5	16	40	72	
Materials & Hardware	6	50	72	60	
Maintenance Practices	7	70	80	60	Mechanical & Electrical Essays
Aerodynamics	8	20	20	20	
Human Factors	9	20	20	20	Essay
Aviation Legislation	10	30	40	40	Essay
Turbine Airplane Aerodynamics, Structures and Systems	11	100	132	N/A	
Gas Turbine Engines	15	60	90	N/A	
Propellers	17	20	30	N/A	

Table 1. Requirements for each type of holder.

When an applicant has passed all of these exams, they continue to the workshop sessions and practical applications. When both are completed they are, subject to approval by their local Aviation Authority, issued their license according to category studied. At this stage, they undertake type specific training for an aircraft and mentored before fully qualified to work independently on an aircraft [4]. The length of study can be three years if run full time. Nowadays, many are work based and part-time where average times to complete are approximately 5-6 years in total. An individual commits extensive time and effort [5].

The order, in which the courses are taken is not specified; there is a natural logic to the order. For example, completing Mathematics and Physics prior to Aerodynamics would be logical; whilst Aviation Legislation could be taken to suit. Typically, the first five subjects are the academic knowledge needed and the remaining aircraft practices and knowledge. In the sector pass rates vary and an average is not a value of use. Some subjects are have lower pass rates, No. 11 that may be accounted for by the highest number of questions. Others, No. 6, frequently have higher pass rates when taken by those with practical experiences.

Figure 1, Aircraft Maintenance in a Hangar.

Multi-Choice Questions

Multi-Choice Question, MCQ, exams are objective orientated assessments to test a specific type of response or fact. There use and acceptance owes much to the US Army in post-war when over a million troops were being assessed and evaluated with a focus to reduce time and workload. It was Wood [6] who took the developed concept to make it accepted and used globally. The use of computer systems to grade faster has only added to its popularity.

MCQ have advantages, with limited supervision and costs being argued as a driving force above the others. It is excellent at establishing comparisons between groups and over time. Disadvantages are varied; for example, one simple numerical mistake will result in a zero when most of the complex part completed successfully [7]. There have been attempts that a wrong answer results in a negative mark awarded and opinion is divided as to its effectiveness. Many argue that it discourages thought and penalizes errors more than other assessments.. These have been applied but nowadays less likely. Those that argue they are not suitable for higher-level assessment give answers justify. Like any assessment, it has a role, and should form part of the overall evaluation. If a student has met the learning objectives then that is the principal task to be clarified and agreed [8]. Regardless if there are better ways, what are the limitations and concerns in this application that are addressed. It is also argued by many that influences are ignored or allowed, e.g., bad handwriting.

EASA Multi-Choice Questions

In all EASA MCQ exams the questions have three possible answers, and each question has 1 minute 15 seconds for completion. This is the time assumed that an average student would take to read, comprehend, read the answers and select the answer. Thus, a 20-question exam will have 25 minutes and a 60-question exam 75 minutes.

A typical exam question is shown below in Table 2, with the corresponding options in answers. This could be a question from the Aerodynamics module.

Question	The centre of gravity of an aircraft can be classified how?
a	It is always behind the centre of lift
b	It is always ahead of the centre of lift
c	It must be in the same position as centre of lift to fly straight and level.

Table 2. A typical MCQ and possible answers.

This particular question is classic in its answer selection[9]. The three answers are designed to make the guessing restrictive to passing and having 75% means a student needs extensive knowledge to reach this hurdle. There are three levels of questions that are asked and can be classified as level 1, 2 and 3. The first is recalling a fact, e.g., what is 4 times 4. The second usually requires to place theory in an application, e.g., if a pilot pushes the control stick forward what direction will the control surface move? The third, more complex, might be, if the voltage drops from Auxiliary Power Unit what is the most likely reason? Where these limit the result is that any student is equally likely to select the correct answer as any other that does not know, regardless of the complexity of the question and level of their knowledge [10].

Statistics of Multi Choice Questions

The probability of a MCQ depends on the number of possible answers set for each question. For the examples used in these exams the three choices result in a probability of guessing correctly as $p = 0.333$ and that of incorrect is $q = 0.666$. Remembering all possible guesses produce probabilities must equal 1 (100%).

Given the probability stated for this design of MCQ a student with no knowledge will score, on average, 33%. We can significantly estimate the variability by probability theory. The average:

$Ave = p \cdot n$, where n – number in the sample (total questions).

Standard deviation, $\sigma = \sqrt{n \cdot p \cdot q}$

Thus, $Ave = 33.3\%$ and if $n = 52$ then $\sigma = 3.4$

Knowing these values and that the 99.73% of variability will be the average $\pm 3\sigma$. The variability of guessing (a good, average or bad) can be determined and is shown in Table 3 below. The table shows the expected average from guessing, σ of that guessing, the range of the population and fially the expected percentage a student could achieve as a result.

Module Title	No.	B1	Average	σ	Range	% range
Mathematics	1	30	10	2.6	2.2 - 17.8	3.3 - 59
Physics	2	52	17.3	3.4	7.1- 27.7	13.6 – 53.2
Electrical Fundamentals	3	52	17.3	3.4	7.1- 27.7	13.6 – 53.2
Electronic Fundamental	4	20	6.66	2.1	0.36 – 12.96	1.8 – 64.8
Digital Techniques	5	40	13.3	3	4.3 – 22.3	10 – 55.7
Materials & Hardware	6	72	23.9	4	11.9 – 35.9	16.5 – 49.8
Maintenance Practices	7	80	26.6	4.2	14 – 39.2	17.5 - 49
Aerodynamics	8	20	6.6	2.1	0.36 – 12.96	1.8 – 64.8
Human Factors	9	20	6.6	2.1	0.36 – 12.96	1.8 – 64.8
Aviation Legislation	10	40	13.12	3	4.2 – 22.3	10 – 55.7
Turbine Airplanes Aerodynamics, Structures and Systems	11	132	50	5.4	33.8 - 76.2	25.6 – 57.7
Gas Turbine Engines	15	90	30	4.45	16.65 - 43.35	1.1 – 48.1
Propellers	17	30	10	2.6	2.2 - 17.8	3.3 - 59

Table 3. Probability of guessing answers.

These values are all based upon probability theory and with a pass mark of 75% it is virtually impossibile to guess to a pass value. A point to note is that the smaller fewer the number of questions set the higher the probability f guessing near to the pass mark. The range in percentages is based on the central limit theorem and explains 99.73% of all possibilities. Referring to Table 1, the Mathematics exam for Category A contains 16 questions. It would be possible to guess 68.5% within the normal range. Furthermore, 10 question predicts guessing could achieve 77.97% and thus a pass. As there are no exams with that few questions it can be statistically argued that the design format is robust and the probability of guess your way to a pass is not impossible, it is statistically very low. These results do show why MCQ with fewer questions do have failure closer to the 75% pass mark than the ones with a much higher number of questions.

In reality, students are unlikely to not revise and are prepared, even if not suitably prepared for all the theory being examined. If the starting point for an assumption is that they know half of the answers, they will only score 50%. This is not true, as incorrect answers are not penalized in any way [11]. The remaining questions a student would guess. If a paper has 100 questions and a student only know 50% then their result will be, on average:

$$50 + (50 \cdot 0.33) = 66.65\ \%\ \text{(Fail)}$$

Range on this guessing will be $56.65 - 76.56\ \%\ (\sigma = 3.33)$

If a student knows 65% of the answers, their result will be, on average:

$65 + (35 \cdot 0.33) = 76.65\ \%$ (Pass)

The implications of this are that the true pass rate is actually less than 65%; actually 63% correct will result in a 75% pass rate being achieved. There are, naturally, nominal variations on these, they do represent the average [12]. Theory suggests this is more likely with MCQ that have fewer questions, but equally possible for all exams.

If a student does fail, they have to wait 90 days before the first re-sit exam can be undertaken, it can be 30 days if remedial and documented study is completed; however, most wait 90 days. The Civil Aviation Authority, CAA, state that the exam bank of questions must have at least enough questions for three unique papers, ideally five. They also state that there must not be more than 20% repeat questions in the repeat exam. Any student revising can study all the questions set previously and be prepared. Given this is an option for a dedicated individual or group of students that revise collectively can be fully prepared for the re-sit exam. With this assumption, that means each student in theory starts with 20% towards their 75% target. Their pass target is now less demanding than before and overall their probability of passing has considerably increased. The statistics of passing and number of questions needed to be known and not guessed will be different. Again, assuming a 100 MCQ, stating with 20% from a previous paper and a student only knows 40% of the remaining questions:

$20 + 40 + (40 \cdot 0.333) = 73.3\ \%$ (Fail)

If they knew 42% of the remaining questions, their result would be:

$20 + 43 + (37 \cdot 0.333) = 75.31\ \%$ (Pass)

This has significant implications for the assessment process if a student can pass by only knowing 43% of the remaining questions, remembering previous questions and guessing the rest. If the amount of repeat questions were only 10% this has the following effect:

$10 + 53 + (37 \cdot 0.33) = 75.3\ \%$ (Pass)

Therefore, the number of questions to know would be still lower than for the first exam. This adds to the reasons why students' success in re-sits are often better. It is not more revision; it is accounted also by the statistics of the exam format used in these MCQ. Consistent guessing correctly with more questions will reduce the success rate and shorter exams have a statistical bias.

Alternative answers

In Table 2 is a typical question and style of answers available. They are designed to remove guessing as a viable option. It is very difficult to derive two possible, but incorrect, answers for all questions. You can state facts to confuse the non-esoterically prepared candidates; however, there is a limit. Missing units, wrong units and scales of units are frequently used for numerical questions. It is not possible for those that require a statement or description. Studies and reviews of answers may have the majority of the correct answers being the longest answer, as it contains

all the detail to clarify. A standard review of answers is needed to ensure incorrect answers are not obvious the longer answers most likely. All other exams in education have independent reviews, the CAA will audit questions, a sample, but this an exam bank of potentially 2000 - 3500 questions for Category B1 alone. Any audit is a snap-shot of the population and cannot be realistically likely to find sufficient trends or styles for correction.

Implications

It is one of the first things students ask, the pass rate, and 75% is relatively high. Consider the M11 exam that may have 3500 pages of notes, a student can think at the start of the course that this could be impossible to revise. Nevertheless, this theory has shown that an exam technique to passing has an influence that is significant in passing on the first attempt and even higher on the second attempt. It can be argued that 20% repeat questions is unlikely on the first repeat. The counter argument is that with only sufficient questions for three original papers this likelihood is high and without intervention as high as 33%. Therefore, 20% is very probable on the first re-sit, and extremely likely of the second re-sit.

To address these imbalances in the procedures the following should be considered and implemented. First, the exam bank needs increasing in numbers and that every year a percentage are replaced with new ones. Secondly, a re-sit paper needs to have a higher pass rate if the number of questions in the bank is not increased and replacements added. Thirdly, questions need to be rigorously approved and validated to prevent those where identifying the answer that is clearly wrong are not allowed to be used. Finally, there is a case for a form of penalty for incorrect answers to minimise guessing an answer.

Conclusions

The authors of this paper has addressed Multi-Choice Questions that are used in the academic training of Aircraft Engineers under the EASA standard. It has reviewed the statistical assumptions and expanded the theory of their principle and applications. It has shown the theory to prevent anyone guessing without knowledge is robust. Where it challenges current thinking is to question the pass rate and how re-sits will theoretically have lower level to meet than the first attempt. These were discussed in detail and recommendations and changes suggested. To conclude, the basis is statistically set and if no changes then future students will be offered a lower expectation to pass on a re-sit than first attempt.

Acknowledgements

The authors would like to thank Mr. R. Pepper, the examinations at LRTT for his inputs and clarification of the CAA examination process. Also to LRTT, in particular, Ian Fitzpatrick for his support in the work and presentation of data.

References

[1] EUROPA – Agencies and other EU bodies – EASA". Europa.eu. Retrieved 2012-10-16.

[2] Regulation of the European Parliament and of The Council". Eur-lex.europa.eu. Retrieved 2012-10-16.

[3] Regulation (EC) No 216/2008 of the European Parliament and of The Council". Eur-lex.europa.uu. Retrieved 2012-10-16.

[4] http://www.eca.europa.eu/Lists/ECADocuments/SR12_15/SR12_15_EN.PDF

[5] Links to National Authorities". EASA. Retrieved 2013-05-16.

[6] Kehoe, Jerard (1995). Writing multiple-choice test items Practical Assessment, Research & Evaluation, 4(9). Retrieved February 12, 2008.

[7] Beckert, L., Wilkinson, T. J., & Sainsbury, R. (2003). A needs-based study and examination skills course improves students' performance Medical Education 37 (5), 424–428. doi:10.1046/j.1365-2923.2003.01499.x

[8] Steven M Downing (2004) Reliability: on the reproducibility of assessment data Medical Education 38 (9), 1006–1012. doi:10.1111/j.1365-2929.2004.01932.x

[9] DePalma, Anthony (1 November 1990). "Revisions Adopted in College Entrance Tests". New York Times. Retrieved 22 August2012.

[10] Bontis, N., Hardie, T., & Serenko, A. (2009). Techniques for assessing skills and knowledge in a business strategy classroom International Journal of Teaching and Case Studies, 2, 2, 162-180.

[11] Leddy, P & Ormrod, J. (2013) Practical Research, 9th Ed., Pearson.

[12] Shavelson, R. (2010) Statistical Reasoning for Behavioral Sciences, 3rd Ed., Allyn & Bacon.